24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

DEVELOPMENT OF A TEST SYSTEM FOR VOICE MONITORING CONTACT SENSOR: PHONATORY SYSTEM SIMULATOR

Federico Casassa

DET – Electronic and Telecommunications Dept., Politecnico di Torino, Turin, IT,
email: federico.casassa@polito.it

Adriano Troia

Metrology for Quality of Life, I.N.Ri.M., Turin, IT,

Alessandro Schiavi

S.T.A.L.T, I.N.Ri.M., Turin, IT.

Contact microphones are the best tool for monitoring the health of the phonatory system in a non-invasive way, to record the voice activity in order to extract reliable parameter useful to diagnose voice diseases like dysphonia, polyps and nodules. After testing different contact microphones on many human subjects arise the need of a proper characterization, since the in vivo tests suffer from lack of reproducibility due to the biological variability and to the absence of a strong reference. For this purpose a phonatory system simulator has been created, an apparatus that can mimic the biological system on which the sensors has to work. Such an apparatus has been validate and four different contact sensor has been tested with it.

Keywords: vocal monitoring, contact microphones characterization.

1. Introduction

Vocal monitoring consists in recording a person's voice activity in order to extract parameter to quantify the voice quality, and detect disorder correlated to the prolonged use of voice. This kind of monitoring is important for those workers categories called voice professionals [1-6] such as teachers, singers and call center operators, who use the voice as a work tool. Vocal holter / vocal accumulator are devices developed for this purpose [7-12]: the voice activity is recorded by means of a contact microphone, which senses the skin vibration at the jugular notch (the base of the neck). The vibration is produced by and correlated to the phonation: the air column in the trachea drives the vocal fold in self-oscillation and creates a pressure signal which is the source of the phonation, but it also produces a vibration in the tissues next to the glottis. The reason why this kind of recording is preferable to a classic acoustic acquisition with a normal microphone is that a contact/vibration sensor is less influenced from the acoustic background noise present in the room where the monitoring takes place, and give more precise information on the actual voice activity of the subject under monitoring. Anyhow, a contact microphone – based vocal holter device needs an air microphone, in order to perform a calibration to relate the intensity of the vibration signal to a sound pressure levels.

The voice descriptor extracted from the recorded vocal signal are frequency and intensity – based parameters like jitter, shimmer, cepstral peak prominence [13-15]. For this reason, the sensor characterization is of primary importance to distinguish between the actual physical phenomena that is being measured and the artifacts introduced by the sensor characteristics. For what concern air microphones the manufacturer present an accurate and precise characterization, but regarding the contact microphone, the specification provided by the manufacturer are often incomplete and unusable for voice monitoring purpose for two main reason: the low-cost devices engineered to be used like laringophones are not conceived for research or diagnostical/medical purposes, and the common vibration transducers are often designed only for mechanical and industrial use. Tests on human subject are the usual way to characterize this kind of low-cost contact sensors [7-12], but there is a lack of standard that can be used to analyze the response of the various transducers in a vocal monitoring framework. Usually this kind of test consist in several vocal tasks (sustained vocal at different pitch and intensity, reading a written passage, free speech) performed by various subjects for each sensor under examination, and the comparison of the parameters obtained from them. However, this kind of tests suffer from lack of reproducibility since the human body is not a stable system on which a proper characterization can be performed [16]. For this reason arise the need of an apparatus that can mimic the biological system on which the sensor has to work, a device that gives two well-know outputs to a source signal: an acoustical signal and a vibration signal on a skin-like material.

The relationship between the vibration at the base of the neck and the acoustical vocal signal is hard to be accurately obtained. Both the signal are related to the source of phonation, i.e. the air that flows through the glottis and drive the vocal folds into a self-oscillating motion, but the response of the two filters (tissue at the base of the neck and vocal tract) are not well reproducible due to various factors: tissue thickness, body fat, sensor position, attachment methods and sensitivity to acoustic stimuli. Furthermore it is hard to understand the features of the signal due to the characteristics of the measurand and the artifacts introduced by the transducers used to acquire the data.

In order to obviate to this problem, a similar system to test the vibration sensors to be used for recording lung sound has been already developed and tested [17, 18]. This device mimic the respiratory system and is able to give a well-known response to a certain stimula. This work deal with a similar apparatus developed on the base of the phonatory system, which acts as a stable generator that can make contact sensors characterization free from reproducibility problems. Moreover, the proposed system allows other characteristic of the sensor to be estimated, that cannot be obtained in vivo such as the load effect of the sensor due to its dimension, shape and weight. The sensitivity to acoustical noise can also be easily investigated using an external sound source, avoiding the problems that have been faced during the in vivo measurements [16].

2. The apparatus

In this section will be provided a description of the human phonatory system and the simulator.

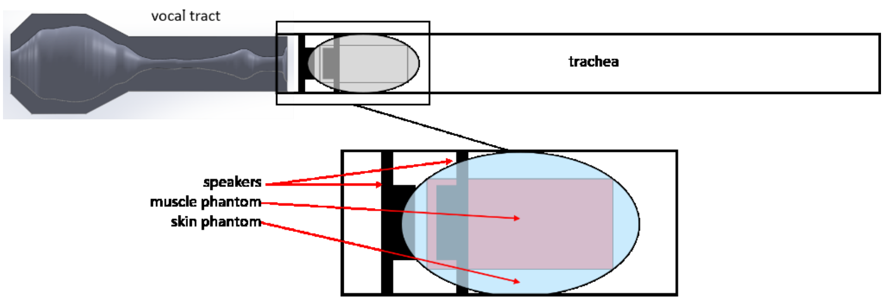

Figure 1: phonatory system simulator scheme, horizontal section (top), detail of the sensing zone (bottom).

2.1 The phonatory system description

.The phonatory system is the source of the voiced sound, and it is made up of the trachea, the larynx and the vocal tract. The larynx is the real source of phonation: it contains the vocal folds, two infolding of tyroarytenoid muscle, and the glottis, the aperture between them. An air column ascends the trachea, which is basically an air channel that convolves in the larynx the air that comes from the lungs, and drives the vocal folds into a self-oscillation motion. In this way the glottis opens and closes itself and creates a pressure wave at the exit of the larynx. This vocal folds oscillation also creates the so called subglottal resonances, pressure standing waves in the trachea [19]. The primary pressure wave, the one which travel from the glottis towards the mouth, passes through the so called "vocal tract", the terminal tract of the pharynx and the oral cavity.

The vocal tract is, from an acoustic point of view, a resonator: it changes the harmonic content of the original pressure wave by filtering some frequencies and amplifying some others, the so called "formant frequencies". This particular frequencies differs in every voiced sound (vowels mainly, and the few voiced consonants like /m/ and /n/) and depends on the oral cavity, lips and tongue disposition: different configuration correspond to different vowels, and each vowel is characterized by their typical formant frequencies. Nowadays the vocal fold activity is usually monitored with the electrostroboscopy, an optical device that allows to see the motion of the vocal folds. In this case, in order to record a signal of the vocal folds movements, an electroglottograph (EGG) has been used: an impedance meter that sense the relative impedance between the vocal folds and is able to record the opening-closing cycle of the glottis. The spectral envelope of this signal is characterized by the greatest amplitude of the fundamental frequency whilst the other harmonic components amplitude decrease as the frequency increase. In the spectral envelope of the acoustical voice signal, the higher harmonics (the so called formant frequencies) are instead stronger than the fundamental frequency, due to the resonant effects of the vocal tract.

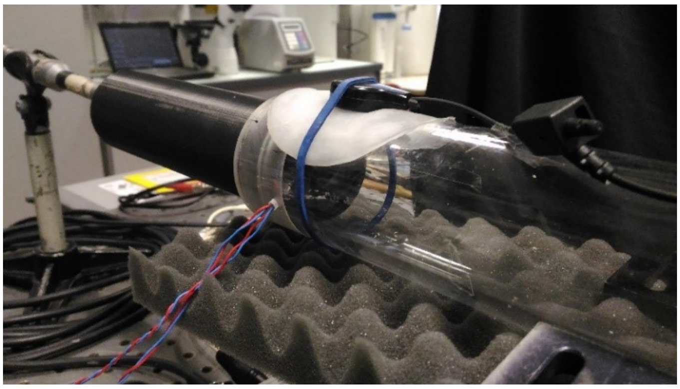

Figure 2: the simulator with the ECM attached on the sensing zone.

2.2 Simulator description

The aim of this project is to develop a system that is able to mimic the vocal apparatus (trachea, glottis/vocal chords, neck tissue and vocal tract) in order to obtain a test system with well-known relations between the vibration of a human-tissue-like material and an acoustic signal. For this reason it is not required to develop an exact mechanical replica of the vocal apparatus like in [20-22], because the goal is the characterization of contact sensors in a stable vocal monitoring framework. The main parts of the vocal apparatus has been recreated as follows, as shown in figure 1:

- trachea: hollow plexiglass tube with a diameter of 5 cm;
- glottis / vocal folds / source signal: two speakers that emit in two different direction;
- vocal tract: a 3D-printed model of Human Vocal Tract;
- neck tissues: tissue-mimicking phantom material (skin and muscles).

The two speakers recreate the glottis activity by emitting two pressure signals: one signal propagates in the plexiglass tube (trachea) and the other one propagates in the vocal-tract simulator.

The plexiglass tube used to simulate the trachea is 55 cm long and it is equipped with a moving removable end, in order to recreate the subglottal resonances by means of the closed/open tube modes. These can be tuned by changing the length of the plexiglass tube, to obtain the right frequencies to mimic the subglottal resonances.

The Vocal tract is a 3D-printed hollow resonator, which is based on a model of the vocal tract while emitting an \a\ vowel proposed by Švec et al. [23]. This resonator allows the formant frequencies from the pressure signal originated from the speaker to be amplified.

As a coupling phantom, a multi component Tissue Mimicking Materials (TMM) has been realized. Two types of muscle/throat simulating tissue have been investigated as TMM: a stiff Gellan Gum based hydrogel containing kieselghur and silicon carbide solid particles, whose detail on preparation and characterization of its acoustic and mechanical properties can be found in [24-25]. Alternatively, a simple latex rubber based TMM, which showed similar properties but with a better time stability, is appeared as ideal to perform several trials required for a complete characterization of the contact sensors. The results presented in this paper are obtained with the latex rubber based TMM. A polyvinyl alcohol based phantom have been prepared with freezing/thaw cycle technique and used as skin simulating tissue. Further details on its preparation can be found in [26].

3. Experimental results

Specific test were carried out to validate the effectiveness of the phonatory system simulator. The aim was to briefly characterize the two filter (glottis-vibration and glottis-voice) and to test a few contact sensors which could be used for voice monitoring: a Midland MIAE38 electret contact microphone (ECM), a Meas-Spec CM-01B piezo film contact microphone (PMIC), a Knowles BU-21771 accelerometer (ACC) and ad a commercial piezo throat microphone designed to be used as an external phone microphone (ARCH). The acoustic signals have been sensed with a Behringer ECM2000 microphone. The signals has been acquired by means of a National Instruments USB 6211 acquisition board connected to a PC, with a 22 kHz sampling frequency. The tube length was 50 cm in the closed-end configuration and 55 cm in the open-end configuration.

3.1 Preliminary tests and validation

A simultaneous recording of EGG, vibration signal at the base of the neck and voice were performed on a human subject; that EGG signal was used to drive the speakers of the simulator to acquire its vibration and acoustic output related to that particular source signal. This test was performed with the simulator in the open-end and closed-end configuration, and the used contact microphone was the PMIC. Hereafter, the spectra of the signals acquired on the simulator (TS) was compared with the spectra of the signals acquired on the human subject (in vivo) (fig. 3).

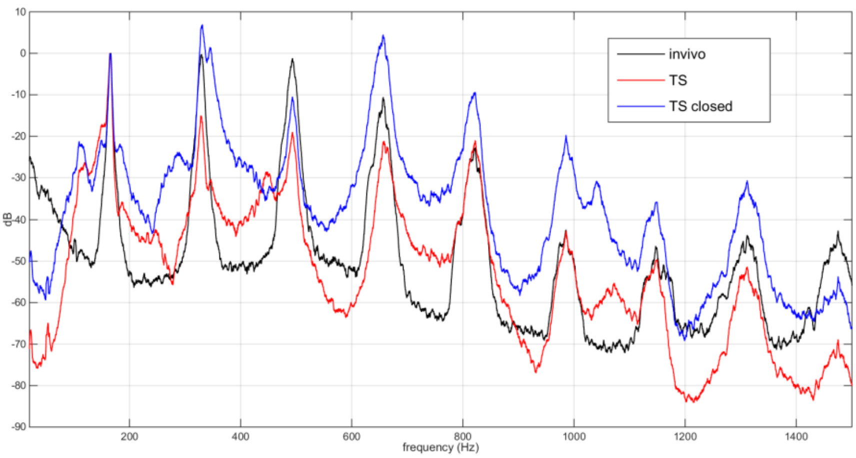

Figure 3: comparison between in vivo and simulator vibration signal spectra.

The same comparison for the acoustical signal is presented in Fig. 4. As it can be seen in the two graphs, the in vivo spectra are quite similar to the TS spectra in the open-end configuration for the first 5 harmonics. The amplitude of the higher harmonics are lower due to the filter effect of the tissue-mimicking material, and to the fact that the vocal tract resonator are made on the base of a vocal tract of a different person among the one who generated the in vivo data.

Anyhow, the resonator amplify the third and the fourth harmonics and create the formant frequencies of the \a\ vowel.

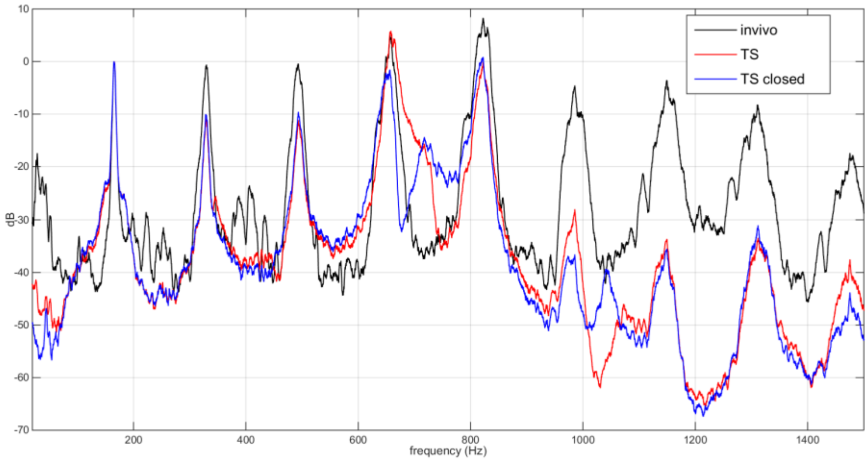

Figure 4: comparison between in vivo and simulator acoustical signal spectra.

3.2 Simulator characterization

The apparatus has been characterized by means of a Laser vibrometer (Polytec CLV-2534), in order to obtain the system vibration response with no mechanical sensor attached to the phantom. The simulator has been placed on the optical bench and the laser was focused on a piece of reflective material placed on the skin-mimicking phantom.

The spectra presented in Fig. 5 refer to the acceleration signal obtained with the vibrometer; in this case the source signal is a frequency sine sweep (10 seconds duration, 100Hz - 1500 Hz). The peaks represent the acoustical resonances of the speakers and the tube in the two configurations (open-end and closed-end). These spectral envelopes define the filter effect of the simulator, which includes the electrical-to-acoustical response of the speakers, the effect of the tube's resonances and antiresonance and the attenuation of the phantoms. The resonance frequencies of the tube in the open-end configuration can be calculated as:

$$f = \frac{nv}{2(L+0{,}8d)}, n = (1,2,3,\ldots) \qquad (1)$$

For the closed-end configuration, the resonance frequencies are:

$$f = \frac{nv}{4(L+0{,}4d)}, n = (1,3,5,\ldots) \qquad (2)$$

where v is the speed of sound in air in (m/s), L is the tube length (m), d is the tube diameter in (m). The peaks correspond to these frequency according to the simulator dimensions.

In order to better underline the vocal tract resonator behaviour and in general the relations between the vibration signal and the acoustic signal, in Fig. 6 the spectra obtained with the EGG source signal, normalized on the magnitude of the first harmonic, are presented. The vibration signal is recorded using the ECM contact microphone. It is noticeable how the contact microphone spectral envelope is quite near to the vibrometer one, with small differences on the 3rd and higher harmonics (the maximum difference is 8 dB). The microphone in air spectral envelope is instead very different from the other, with the 3rd and 4th harmonics amplitude higher than the ones in the other two envelopes. The differences between the microphone in air and the contact sensor in closed-end configuration is 2 dB, 16 dB and 32 dB for the 2nd, 3rd and 4th harmonic respectively, and 12 dB, 35 dB and 34 dB for the same harmonics in open-end configuration

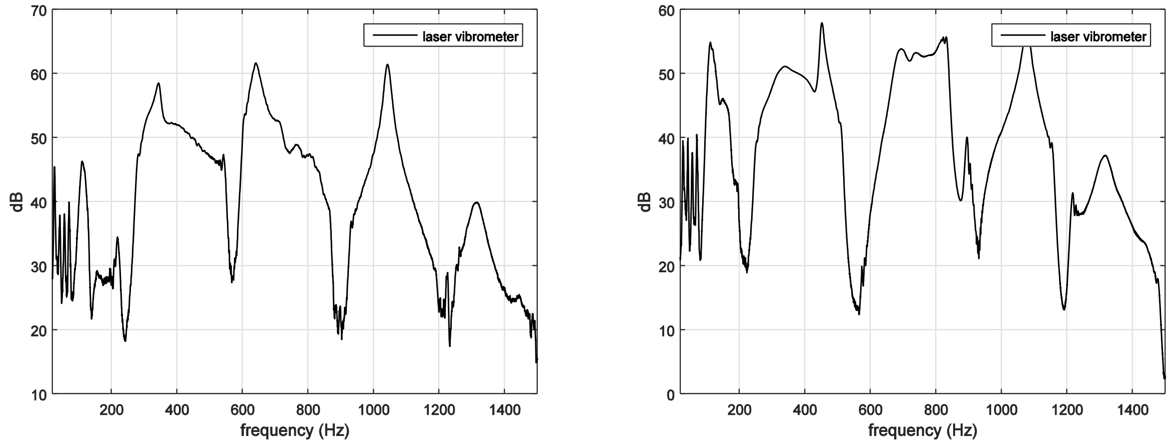

Figure 5: Vibrometer spectral envelopes obtained with a frequency sine sweep source signal, closed-end (left) and open-end (right) configuration.

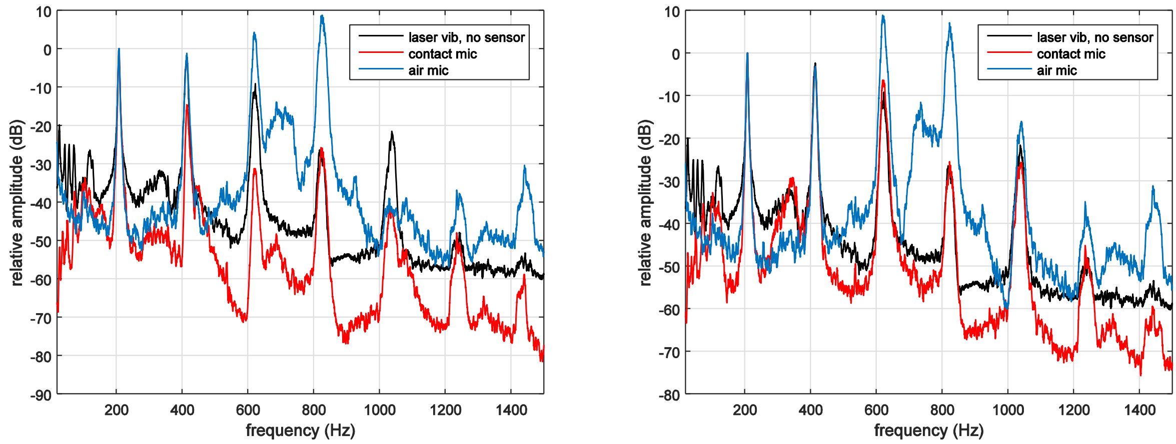

Figure 6: Vibrometer, contact sensor and microphone in air spectral envelopes obtained on the simulator with EGG source signal, closed-end (left) and open-end (right) configuration.

3.3 Sensor characterization

Further experiments were carried out with the aim of underline the responses of different sensors. All the four previously mentioned transducers were test on the TS, with a frequency sweep (15 seconds duration, 40 Hz – 5000 Hz) as source signal, and the spectra obtained are presented in Fig 7. The comparison reveals the differences in the sensors responses, and confirms the necessity of a standard to be used to obtain a proper characterization.

4. Conclusions

The tests carried out shows that the simulator mimics the behaviour of the phonatory system in an efficient way for what concern the initial purpose of the work: testing contact sensors to be used in a voice monitoring framework. The source signal is well transmitted to the phantom’s surface, and the vocal tract resonator amplify the high harmonics of the source signal and create the vowel formants. The system allows to test the contact sensors with a known reference. Hereafter, an artificial EGG-like signal will be created, to be used as a source signal. The last characterization point out the differences between sensors, and this outcome is the evidence that different sensors have a different response that must be evaluated and included in the vocal parameter evaluation.

Next step will be to embed a piezo film stripe within the muscle-mimicking phantom material to estimate the load effect of the sensor.

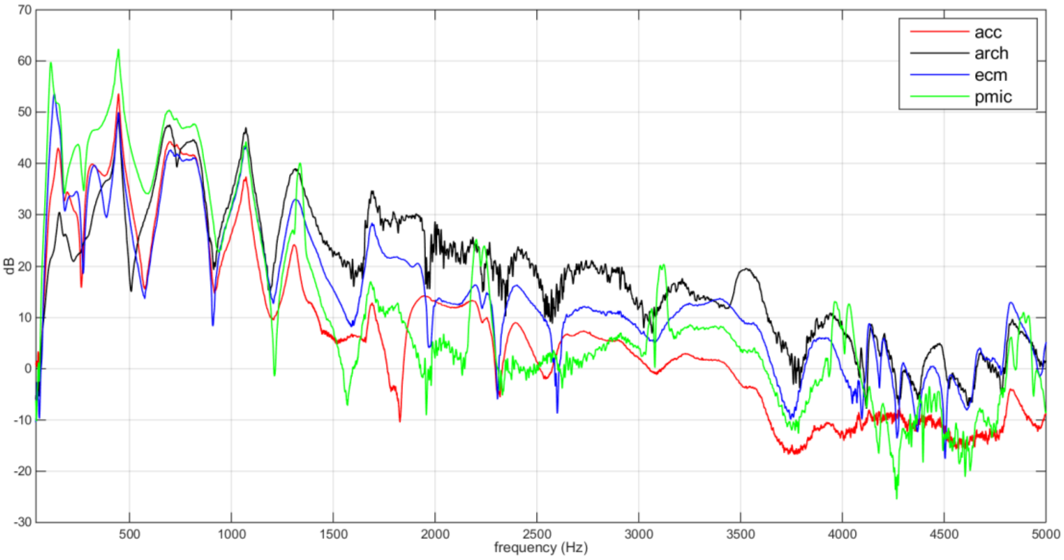

Figure 7: Comparison between 4 different contact sensors response obtained on the TS.

REFERENCES

1 D.D. Metha, M. Zañartu, S. W. Feng, H. A. Cheyne, R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform", IEEE Trans. On Biomedical Engineering, vol. 59, n. 11, pp. 3090-3096, Nov. 2012

2 G. J. Švec, I. R. Titze, P. S. Popolo, "Estimation os sound pressure levels of voiced speech from skin acceleration of the neck", J. Acoust. Soc. Am, vol. 117, no. 3 pt. 1, pp 1386-1394, March 2005

3 M. Ghassemi, J. H. Van Stand, D. D. Metha, M. Zañartu, H. A. Cheyne, R. E. Hillman, J. V. Guttag, "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: initial results for vocal fold nodules", IEEE Trans. on biomedical engineergin, vol. 61, no. 6, June 2014, pp. 1668-1675

4 P. Bottalico, A. Astolfi, "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms", Journal of the Acoustical Society of America, vol. 131 no. 4, pp. 2817- 2827, 2012.

5 A. Astolfi, p: Bottalico,A. Accornero, M. Garzaro, J. Nadalin, C. Giordano, "Relationship between vocal doses and voice disorders on primary school teachers", Proc. Euronoise Conference, Prague, Czech Republic, pp. 55-60, 2012.

6 M. Wirebrand, "Real-time monitoring of voice characteristics using accelerometer and microphone measurements" [Online.] available: http://liu.diva-portal.org/smash/get/diva2:432959/FULLTEXT01.pdf

7 H.A. Cheyne, H.M. Hanson, R.P. Genereux, K.N. Stevens, R.E. Hillman, "Development and Testing of a Portable Vocal Accumulator", Journal of Speech Language and Hearing Research, vol. 46, pp. 1457-1467, 2003.

8 P.S. Popolo, J.C. Švec, I.R. Titze, "Adaptation of a Pocket PC for Use as a Wearable Voice Dosimeter", Journal of Speech Language and Hearing Research, vol. 48, pp. 780-791, 2005.

9 A. Carullo, A., A. Vallan, A. Astolfi "Design Issues for a Portable Vocal Analyzer", IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 5, pp 1084-1093, 2013.

10 A. Carullo, A. Vallan, A. Astolfi, "A Low-Cost Platform for Voice Monitoring", Proceedings of I2MTC Conference, Minneapolis, MN (USA), May 6-9, 2013

11 A. Carullo, A. Vallan, A. Astolfi, L. Pavese, G.E. Puglisi, “Validation of Calibration Procedures and Uncertainty Estimation of Contact-Microphone Based Vocal”, Measurement, vol. 74, pp. 130-142, Oct. 2015.

12 M. Zañartu, C.C. Ho, S. S. Kraman., H. Pasterkamp, J. E. Huber, G.R. Wodicka, “Air-Borne and Tissue-Borne Sensitivities of Bioacoustic Sensors Used on the Skin Surface”, IEEE Tr. on Biomedical engineergin, vol. 56, no. 2, pp.443- 451, 2009.

13 Y.D. Heman-Ackah, D.D. Michael, G.S. Goding, “The relationship between cepstral peak prominence and selected parameters of dysphonia”, J Voice,vol.16, pp. 20–27, 2002.

14 Y. Maryn, M. De Bodt, N. Roy, “The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders,” J. Commun. Disord.,vol. 43, pp.161–174, 2010.

15 S.N. Awan, N. Roy, M.E. Jette, G.S. Meltzner, R.E. Hillman, “Quantifying dysphonia severity using a sepctral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V,” Clin. Linguist. Phon., vol.24, pp.742–758, 2010.

16 A. Carullo, F. Casassa, A. Castellana, A. Astolfi, L. Pavese, G. E. Puglisi, “Performance comparison for different contact-microphones used for voice mointoring”, 22nd Interntion congress on sound and vibration, Florence (ITA), June 12-16, 2015.

17 S.S. Kraman, G. A. Pessler, H. Pasterkamp, G. R. Wodicka, “Desing, construction and evaluation of bioacoustic transducer testing (BATT) system for respiratory sound”, IEEE Trans. on biomedical engineering, vol. 53, no. 8, pp. 1711-1715, Aug. 2015.

18 S.S. Kraman, G. R. Wodicka, G. A. Pessler, H. Pasterkamp, “Comparson of lung sound transucers using a bioacoustic transducers testing system”, J. Appl. Physiol., vol. 101, pp. 469-476, Aug. 2006.

19 S. F. Austin, I. R. Titze, “The effect of subglottal resonances upon vocal fold vibration”, Journal of voice, vol. 11, iss. 4, pp. 391-402, December 1997.

20 J. Horacek, V. Uruba, V. Radolf, J. Vesely, V. Bula, “Airflow visualization in a model of human glottis near the self-oscillating vocal folds model”, J. Apllied and computational mechanics, vol. 5, pp 21-28, 2011.

21 L. P. Fulcher, R. C. Sherer, T. Powell, “Pressure distribuition in a static physical model of the uniform glottis: entrance and exit coefficients”, J. Acoust. Soc. Am., vol. 129, no. 3, March 2001.

22 G. C. J. Hofmans, G. Groot, M. Ranucci, G. Graziani, A. Hirshberg, “unsteady flow through in-vitro model of the glotts”, J. coust. Soc. Am, vol. 113, no. 3, march 2003.

23 T. Vampola, J. Horacek, J. G. Švec, “FE modeling of human vocal tract acoustics. Part 1: production of czech vowels”, Acta acustica united with Acustica, vol. 94, pp 433-447, 2008.

24 A. Schiavi, R. Cuccaro, A. Troia, “Strain-rate and temperature dependent material properties of Agar and Gellan Gum used in biomedical applications”, Journal of the mechanical behavior of biomedical materials, vol. 53, p. 119-130, 2015.

25 R. Cuccaro, C. Musacchio, PA. Giuliano Albo, A. Troia, S. Lago, “Acoustical characterization of polysaccharide polymers tissue-mimicking materials”, Ultrasonics, vol. 56, p. 210-219, 2014.

26 K.J.M. Surry, H.J.B. Austin, A. Fenster, T.M. Peters, “Poly(vinyl-alcohol) cryogel phantoms for use in ultrasound and MR imaging”, Physics in Medicine and Biology, vol. 49, no. 24, pp. 5529-5546, 2004.